Classification of Mars McMurdo panorama images using machine learning techniques

Shang, Changjing; Barnes, David Preston

Publication date:
2011

Citation for published version (APA):
Shang, C., & Barnes, D. P. (2011). *Classification of Mars McMurdo panorama images using machine learning techniques*. http://hdl.handle.net/2160/7628

CLASSIFICATION OF MARS MCMURDO PANORAMA IMAGES USING MACHINE LEARNING TECHNIQUES

Changjing Shang and Dave Barnes

Dept. of Computer Science, Aberystwyth University, UK
email: {cns, dpb}@aber.ac.uk

ABSTRACT

This paper presents a novel application of advanced machine learning techniques for Mars terrain image classification. Fuzzy-rough feature selection (FRFS) is employed in conjunction with Support Vector Machines (SVMs) to construct image classifiers. These techniques are for the first time, integrated to address problems in space engineering where the images are of many classes and large-scale. The use of FRFS allows the induction of low-dimensionality feature sets from feature patterns of a much higher dimensionality. Experimental results demonstrate that FRFS helps to enhance the efficacy of the conventional classifiers. The resultant SVM-based classifiers which utilise FRFS-selected features generally outperform K-Nearest Neighbours and Decision Tree based classifiers and those which use PCA-returned features.

1. INTRODUCTION

Automated and accurate analysis of Mars images obtained by the front-line Panoramic Camera (Pancam) instruments [1, 2] is an important task, especially for surveying places (e.g. for geologic cues) in Mars [3, 4, 5, 6]. A key element of analysing Pancam terrain images is to detect rocks and other objects captured in such images. However, such objects on Mars exhibit diverse morphologies, colours and textures. They are often covered in dust, grouped into self-occluding piles or partially embedded in the terrain [2]. Also, Mars terrain images vary significantly in terms of intensity, scale and rotation, and are blurred with measurement and transmission noise. These factors make Martian image classification a challenging problem [3, 5, 6].

One critical step to successfully build an image classifier is to extract and use informative features from given images [6, 7, 8]. To capture the essential characteristics of such images, many features may have to be extracted without explicit prior knowledge of what properties might best represent the underlying scene reflected by the original image. However, generating more features increases the computational complexity, especially in light of on-board processing of Mars images where demand for computational memory and processing time must be minimised, despite the nowadays generally available and relatively cheap computer power. Besides, not all such features may be useful to perform classification [6, 7, 9, 10]. Due to measurement noise the use of extra features may even reduce the overall representational potential of the feature set and hence, the classification accuracy . Thus, it is often necessary to employ a method that can determine the most significant features, based on sample measurements, to simplify the classification process, while ensuring high classification performance. Recently, there have been significant advances in developing methodologies that are capable of minimising feature subsets in a noisy environment. In particular, a resounding amount of research utilises fuzzy and rough sets [9]. Amongst them is the fuzzy-rough feature selection (FRFS) algorithm [11] that has been shown to be a highly useful technique by which discrete or real-valued noisy data (or a mixture of both) can be effectively reduced, without the need for user-supplied information.

Inspired by this observation, this paper presents an integrated approach for performing large-scale Mars image classification, by exploiting the potential of advanced classification and feature selection techniques. In particular, Support Vector Machines (SVMs) [12] are employed for image classification. This is due to the recognition of their high generalisation performance in complex data sets [12, 13]. FRFS is utilised to ensure that classification is carried out with a subset of original features only.

The resulting integrated approach helps to improve the effectiveness and efficiency of SVM-based image classifiers. This is because only those informative features are required to be generated in performing actual classification, minimising both the feature measurement noise and the computational complexity (of both feature extraction and feature pattern-based classification). Systematic experimental studies are carried out in comparison with the use of classical classification techniques (e.g. Decision

Trees [14], and K-Nearest Neighbours [15]), with features selected by FRAS or conventional dimensionality reduction methods (e.g. PCA [15]). These results show that the proposed approach entails rapid and accurate learning of classifiers. This is of great importance to on-board image classification in future Mars rover missions. This is because flight projects demand least memory requirement and simplest computation possible.

The rest of this paper is organized as follows. Section 2 introduces the Mars terrain images under investigation. Sections 3, 4 and 5 outline the key component techniques used in this work, including feature extraction, (fuzzy-rough) feature selection and supoort vector machine based feature pattern classifiers. Section 6 shows the experimental results, supported by comparative studies. The paper is concluded in Section 7.

2. *MCMURDO* PANORAMA IMAGE

Although the approach taken in this research is general, the present work concentrates on the classification of the 360-degree view *McMurdo* panorama image. This (composed) image is obtained from the panoramic camera on NASA's Mars Exploration Rover Spirit and presented in approximately true color [2], consisting of 1,449 Pancam images and representing a raw data volume of nearly 500 megabytes. Such an image reveals a tremendous amount of detail in part of Spirit's surroundings, including many dark, porous-textured volcanic, brighter and smoother-looking rocks, sand ripple, and gravel (mixture of small stones and sand).

Fig. 1 shows the most part of the original *McMurdo* image (of a size 20480×4124). This image, excluding the areas occupied by the instruments and their black shadows, is used for the work here, involving eight major image classes which are of practical significance. These image classes are listed in Table 1 and illustrated in Fig. 2. The ultimate task of this research is to develop an image classifier that can detect and recognise different class regions within a given image.

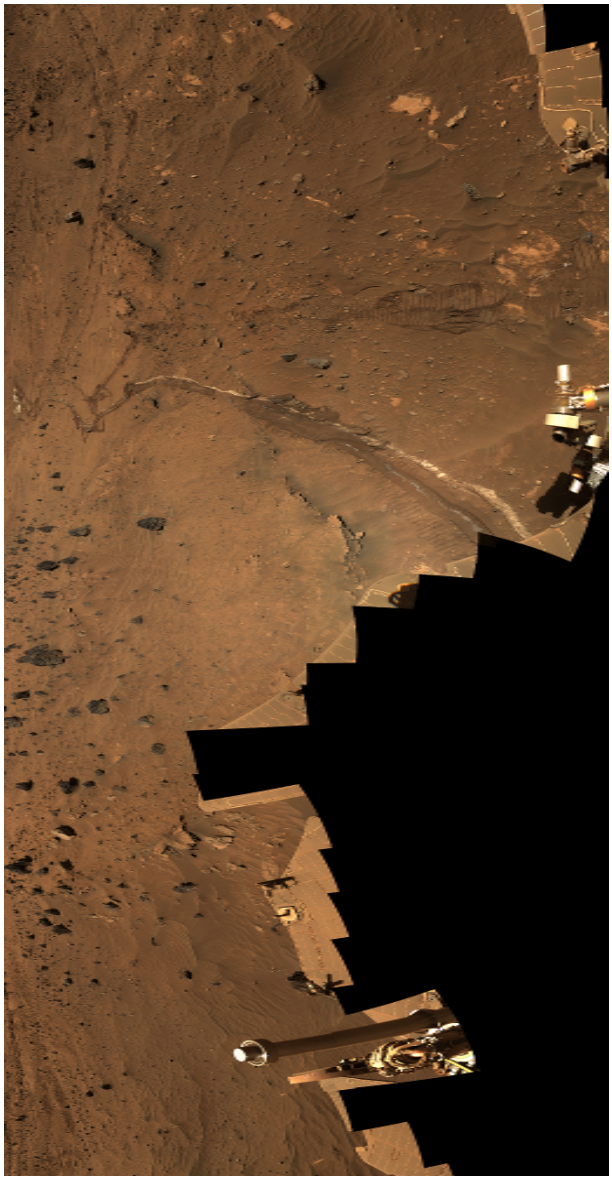

Fig. 1. Mars McMurdo panorama image.

3. FEATURE EXTRACTION

A variety of techniques may be used to capture and represent the underlying characteristics of a given image [7, 16]. In this work, low-level feature extraction approaches are employed. In particular, local colour histograms and the first and second order colour statistics [13, 16] are exploited to produce a feature vector for each individual pixel. Such features are effective in depicting the underlying image characteristics and are efficient to compute. Also, the resulting features are robust to image translation and rotation, thereby potentially suitable for classification of Mars images [6, 13].

3.1. Colour Statistics-Based Features

Images originally given in the RGB (Red, Green and Blue) colour space are first transformed to those in the HSV (Hue, Saturation and Value) space [16]. These spaces are in bijection with one another, and the HSV colour space is widely used in the literature. By computing the first order (mean) and the second order (standard deviation, denoted by STD) colour statistics with respect to each of the R, G, B, H, S and V channels, twelve features can be generated per pixel,

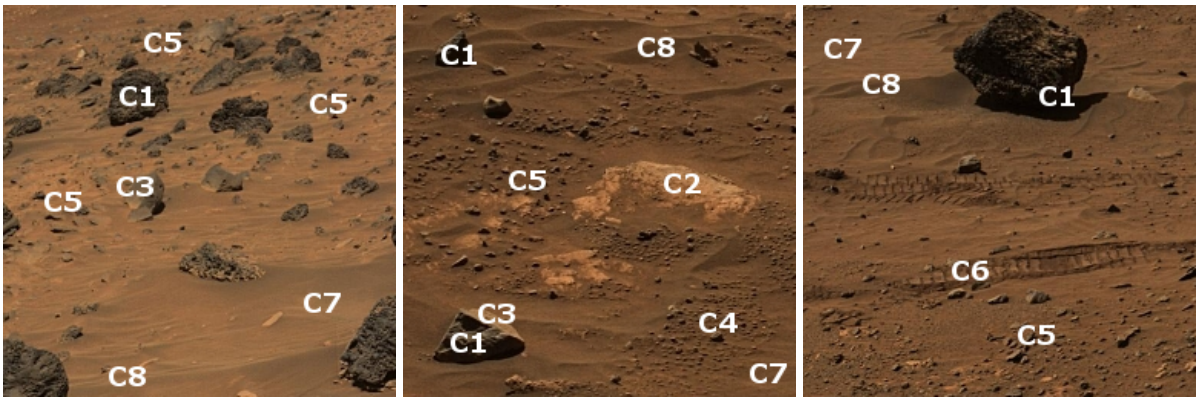

Fig. 2. Image classes as described in Table 1

Class Description	Label
Rock-1 (textured/smoothed dark rock or shadows)	C1
Rock-2 (smoothed bedding orange colored rock)	C2
Rock-3 (graysmoothed rock)	C3
Gravel-1 (mixed with small stone and sand)	C4
Gravel-2 (mixed with small black/orange colored rock and sand)	C5
Rover tracks	C6
Sand	C7
Sand ripple	C8

Table 1. Image classes and class labels

from a certain neighbourhood of that pixel. For presentational simplicity, the resulting features are hereafter denoted as $MEAN_X$ and STD_X, $X \in \{$R, G, B, H, S, V$\}$, representing the first and second order statistics per colour space channel, respectively.

3.2. Local Histogram-Based Features

As the name indicates, such features are measured off the histograms computed from local regions of a given image [17]. In the present context, a histogram is a summary graph showing a count of grey levels falling in a number of resolution ranges (called bins), within a predefined neighbourhood. For a certain pixel, a set of histogram features $X_{hi}, i = 1, 2, ..., B$, where $X \in \{H, S, V\}$, are calculated (within the given neighbourhood), with respect to a particular bin size B (i.e. number of bins), regarding the H, S, and V colour channels. Thus, feature X_{hi} represents the normalised frequency of the colour histogram in bin i. Here, for simplicity, individual bin widths are set equally, and the neighbourhood size is set to the same as that used in the above colour feature extraction. The bin size B is empirically set to 8 in this work.

4. FUZZY-ROUGH FEATURE SELECTION

The theoretical foundation of Fuzzy-Rough Feature Selection (FRFS) algorithm is outlined below; further details can be found in [9, 11].

Let U be the set of pixels within a given image, P be a subset of features, and D be the set of all possible image classes of interest. The concept of fuzzy-rough dependency measure, of D upon P (which FRFS is based on), is defined by [9]:

$$\gamma_P(D) = \frac{\sum_{x \in U} \mu_{POS_{R_P}(D)}(x)}{|U|} \tag{1}$$

where

$$\mu_{POS_{R_P}(D)}(x) = \sup_{X \in U/D} \mu_{\underline{R_P} X}(x) \tag{2}$$

$$\mu_{\underline{R_P} X}(x) = \inf_{y \in U} I(\mu_{R_P}(x, y), \mu_X(y)) \tag{3}$$

and U/D denotes the (equivalence class) partition of the image (i.e. pixel set) with respect to D, and I is a fuzzy implicator and T a t-norm [9]. R_P is a fuzzy similarity relation induced by the feature subset P:

$$\mu_{R_P}(x, y) = T_{A \in P}\{\mu_{R_{\{A\}}}(x, y)\} \tag{4}$$

Here, $\mu_{R_{\{A\}}}(x, y)$ represents the degree to which pixels x and y are deemed similar with regard to feature A. It may be defined in many ways, but in this work, the following commonly used similarity relation [11] is adopted:

$$\mu_{R_{\{A\}}}(x, y) = 1 - \frac{|A(x) - A(y)|}{A_{max} - A_{min}} \tag{5}$$

where $A(x)$ and $A(y)$ stand for the value of feature $A \in P$ of pixel x and that of y, respectively, and A_{max} and A_{min} are the maximum and minimum values of feature A.

FRFS works by employing the above dependency measure to choose which features to add to the subset of the current best features through a greedy hill-climbing process. It terminates when the addition of any remaining feature does not increase the dependency. Fig. 4 outlines this feature selection algorithm; what is returned by this algorithm is a subset of features selected from the full set of original features. Note that as the fuzzy-rough dependency measure is nonmonotonic, it is possible that the hill-climbing search terminates having reached only a local optimum.

C: the set of all original features;
D: the set of possible image classes.
(1)$R \leftarrow \{\}$, $\gamma_{best} = 0$
(2) **do**
(3) $T \leftarrow R$, $\gamma_{prev} \leftarrow \gamma_{best}$
(4) **foreach** $a \in (C - R)$
(5) **if** $\gamma_{R \cup \{a\}}(D) > \gamma_T(D)$
(6) $T \leftarrow R \cup \{a\}$, $\gamma_{best} \leftarrow \gamma_T D$
(7) $R \leftarrow T$
(8) **until** $\gamma_{best} == \gamma_{prev}$
(9) **return** R

Fig. 3. The FRFS Algorithm

5. SVM-BASED CLASSIFIERS

Support Vector Machines (SVMs) [12] are herein used to perform image classification, by mapping input feature vectors onto the underlying image class labels. Such a classifier seeks to find the optimal separating hyperplane among different classes by focusing on those training points (named support vectors), which are placed at the edge of the underlying feature vectors and whose removal would change the solution to be found.

More formally, SMVs construct a hyperplane in a space of a dimensionality higher than that of the original, which is then used for classification (or for other tasks such as regression and prediction). The underlying intuition is that by mapping the original data space into a much higher-dimensional space, the class separation between data points will become easier in that space. SVMs use a specific mapping such that the cross products of data points in the larger space are defined in terms of a kernel function [18] which is selected to suit the given problem. In so doing, the cross products may be computed in terms of the variables in the original space, thereby minimising computational effort. In particular, a hyperplane in the higher dimensional space is defined as the set of points whose inner product with any vector in that space is constant. A good hyperplane is learned over a training process such that the resulting hyperplane has the largest distance to the nearest training data points of any given class. This is in order to increase the discriminative power of the trained classifier.

In the following, Radial Basis function (RBF) kernel is adopted to implement the SVM-based classifiers, and the sequential minimal optimisation algorithm of [19] is used to train the SVMs. Detailed SVM learning mechanism is omitted, but can be found in the literature (e.g. [12, 19]).

In order to increase the efficacy of the SVM classifiers, FRFS is used to rank the extracted features and hence, to select those most informative during the training phase. This is of practical significance as for on-board application, classifiers are expected to be built with mature technologies (rather than totally new mechanisms that have limited experimental data). SVMs are proven high-performance classifiers, but they rely on quality input features. Adding SVMs with FRFS-based feature selection helps to improve the quality of their input. For learning such classifiers, a set of training data is selected from the typical parts (see Fig. 2) of the *McMurdo* image, with each pixel represented by a feature vector which is manually assigned an underlying class label.

6. EXPERIMENTAL RESULTS

From the *McMurdo* image of Fig. 1, a set of 270 subdivided non-overlap images with a size of 512×512 each are used to perform this experiment. 1492 pixel points are selected from 28 of these images for training and verification. Each of the pixels is labeled with an identified class index (i.e. one of the eight image classes: Rock-1, Rock-2, Rock-3, Gravel-1, Gravel-2, Rover tracks, Sand and Sand ripple as listed in Table 1). The rest of all these images are used as unseen data for classification. Each training pixel is represented by a pattern vector of 36 features (see Section 3). The size of a neighbourhood window used for extracting features is set to 15×15. Of course, the actual classification process only uses subsets of selected features.

For comparison purposes, the commonly used K-Nearest Neighbours (KNN) [15] and Decision Trees (DT) [14] classifiers are also employed. The performance of each classifier is measured using classification accuracy, with ten-fold cross validation. For easy cross-referencing, Table 2 lists the reference numbers of the original features that may be extracted, where $1 \leq i \leq 8$ (the empirically chosen bin size). The SVM penalty parameter is set to 100, with standard Gaussian Radial Basis function (RBF) employed. Note that in the following, for KNN classification, the results are first obtained with K set to 1, 3, 5, 8, and 10. Those classifiers which have the highest accuracy, with respect to a given feature pattern dimensionality

No.	Meaning	No.	Meaning	No.	Meaning
1	$MEAN_R$	2	STD_R	3	$MEAN_G$
4	STD_G	5	$MEAN_B$	6	STD_B
7	$MEAN_R$	8	STD_R	9	$MEAN_G$
10	STD_G	11	$MEAN_B$	12	STD_B
13-20	H_{hi}	21-28	S_{hi}	29-36	V_{hi}

Table 2. Feature meaning and reference number

and a certain number of nearest neighbours, are then taken to run for performance comparison.

6.1. Use of Full Original Features

This subsection shows that, at least, the use of a selected subset of features does not significantly reduce the classification accuracy as compared to the use of the full set of original features. For this problem, FRFS returns 8 features out of the original thirty-six (whose references are listed in Table 2). The selected features are: STD_R, $MEAN_G$, STD_B, $MEAN_H$, STD_S, H_{h4}, H_{h5}, S_{h1} (see section 3 for their underlying meaning), with the reference numbers being $2, 3, 6, 7, 10, 16, 17,$ and 26 respectively. This indicates a dimensionality reduction rate of 78%. Table 4 lists the correct classification rates produced by the SVM, DTREE and KNN classifiers, all with 10-fold-cross-validation, where the number of the nearest neighbours K used by these KNN classifiers are also provided (in the first column). Clearly, the classification accuracy of using the eight FRFS-selected features is higher than that of using the thirty-six original features for SVM classifiers (92.63% vs. 92.06%). For KNN and DTREE classifiers, the accuracy resulting from using the eight FRFS-selected features remains very close to that from using the full set of original features (86.12% vs. 86.39% for KNN, and 79.01% vs. 80.02% for DTREE). Overall, the combined use of SVM and FRFS techniques offers the best performance, with a classification rate of 92.63%. This is indicative of the potential of FRFS in reducing not only redundant feature measurements but also the noise associated with such measurements, improving both effectiveness and efficiency of the classification process.

Classifier	Set	Dim.	Feature No	Rate
SVM	FRFS	8	$2, 3, 6, 7, 10, 16, 17, 26$	92.63%
SVM	Full	36	$1, 2, ..., 35, 36$	92.06%
KNN(K8)	FRFS	8	$2, 3, 6, 7, 10, 16, 17, 26$	86.12%
KNN(K5)	Full	36	$1, 2, ..., 35, 36$	86.39%
DTREE	FRFS	8	$2, 3, 6, 7, 10, 16, 17, 26$	79.01%
DTREE	Full	36	$1, 2, ..., 35, 36$	80.02%

Table 3. FRFS-selected vs full original features

6.2. Use of Randomly Selected Features

The above comparison ensured that little information loss is incurred due to fuzzy-rough feature reduction. The question now is whether any other feature sets of a dimensionality 8 would perform similarly as those identified via fuzzy-rough selection. To avoid a biased answer to this, without resorting to exhaustive computation, 30 sets of eight features randomly chosen were used to see what classification results might be achieved.

Figure 3 shows the correct classification rates of the corresponding 30 classifiers, along with the classification rates of the three classifiers that each use the eight FRFS-selected features. Table 4 further summarises these results, where the second, third and forth columns present the worst, average and best classification rates with the corresponding feature sets used. The average rates of the classifiers that each employ eight randomly selected features are only 76.44% for SVM, 73.15% for KNN, and 69.89% for DTREE, far lower than those attained by their counterparts which utilise the FRFS-returned features (of the same dimensionality). The best result that a randomly selected feature set achieves is 84.58% (with features $3, 6, 9, 13, 15, 18, 24, 27$) when used by an SVM classifier. This is much lower than that of using FRFS-selected features (i.e. $2, 3, 6, 7, 10, 16, 17, 26$). This implies that randomly selected features led to important information loss in the course of feature reduction; this is not the case for the FRFS approach.

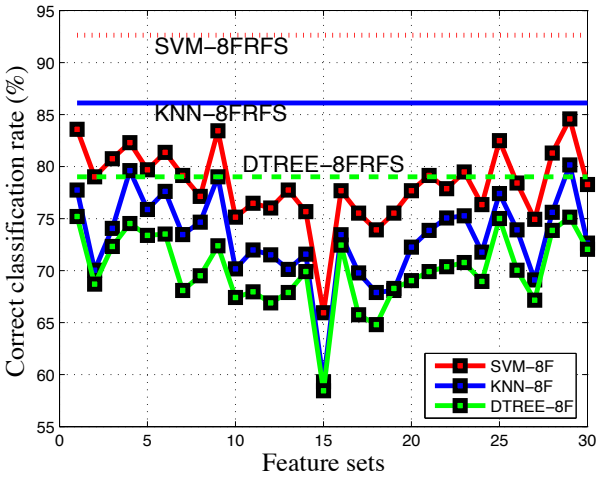

Fig. 3. FRFS vs randomly selected features.

6.3. Use of PCA-Returned Features

Experimentation carried out in this study aims at examining the classifier performance while using differ-

CLF	Min.	Ave.	Max.	FRFS
SVM	65.95%	76.44%	84.58%	92.63%
KNN	59.31%	73.15%	80.16%	86.12%
DTREE	58.44%	69.89%	75.13%	79.01%

Table 4. FRFS vs randomly selected original feature sets of the same dimensionality (8)

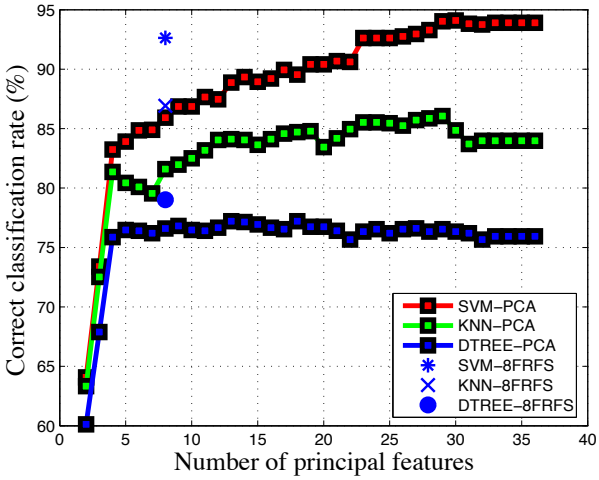

Fig. 4. Performance of KNN, DTREE and SVM vs. the number of principal components.

ent dimensionality reduction techniques. In particular, classifiers that are aided with FRFS are systematically compared to those supported by the use of PCA [15] which is arguably one of the most popular methods for dimensionality reduction, it is adopted here as the benchmark for comparison. Fig. 4 shows the classification results of the SVM, KNN, and DTREE classifiers using a different number of principal features. For easy comparison, the results of the KNNs, DTREE and SVM which use 8 FRFS-selected features are also included, which are represented by $\times$, $\bullet$ and $*$, respectively.

These results demonstrate that the three classifiers which use FRFS-selected features have a substantially higher classification accuracy than their counterparts which use a subset of PCA-returned features of the same dimensionality (92.63% vs. 83.58% for SVM, 86.12% vs. 80.22% for KNN, and 79.01% vs. 73.58% for DTREE,). This is achieved via a considerably simpler computation, due to the substantial reduction of the complexity in the input patterns. The figure also systematically presents the other cases where PCA-aided (SVM, KNN and DTREE) classifiers each employ a feature subset of a different dimensionality. However, these classifiers still generally underperform than the corresponding FRFS-aided ones, whether they are implemented using SVM or KNN. This situation only changes when almost all the PCA-returned features are used where the corresponding SVM classifiers may perform similarly or slightly better (if the number of principal components is larger than 25). Yet, this is at the expense of requiring many more feature measurements and much more complex classifier structures. Besides, PCA alters the underlying semantics of the features during its transformation process. Those features returned by PCA are not the original ones, but their linear combinations.

6.4. Classified and Segmented Images

The ultimate task of this research is to classify Mars panoramic camera images and to detect different objects or regions in such images. The SVM which employs the 8 FRFS-selected features, and which was trained by the given 1492 labeled feature patterns, is herein taken to accomplish the classification of the entire image of Fig. 1 (other than those excluded regions as indicated previously). As an illustration, five classified images are shown in Fig. 5, numbered by (a), (b), (c), (d) and (e) respectively, where eight different colours represent the eight image types (Rock-1, Rock-2, Rock-3, Gravel-1, Gravel-2, Rover tracks, Sand and Sand ripple as listed in Table 1). From this, boundaries between different class regions are identified and marked with white lines, resulting in the segmented images also given in Fig. 5, correspondingly numbered by (f), (g), (h), (i) and (j).

From these results, it can be seen that regions belonging to the eight image classes vary in terms of their size, rotation, colour, contrast, shapes, and texture. For human eyes it can be difficult to identify boundaries between certain image regions, such as those between sand and gravel, and those between rocks and sand. However, the classifier is able to perform under such circumstances, showing its robustness to image variations. This indicates that the small subset of features selected by FRFS indeed convey the most useful information of the original. Note that classification errors mainly occur within regions representing sand and gravel. This may be expected since gravel is itself a mixture of sand and small stones.

7. CONCLUSION

This paper has presented a study on Mars terrain image classification, supported by advanced machine learning techniques. For the first time, fuzzy-rough feature selection has been adopted in conjunction with Support Vector Machines to help solve problems in space engineering. Unlike transformation-based dimensionality reduction techniques, this approach

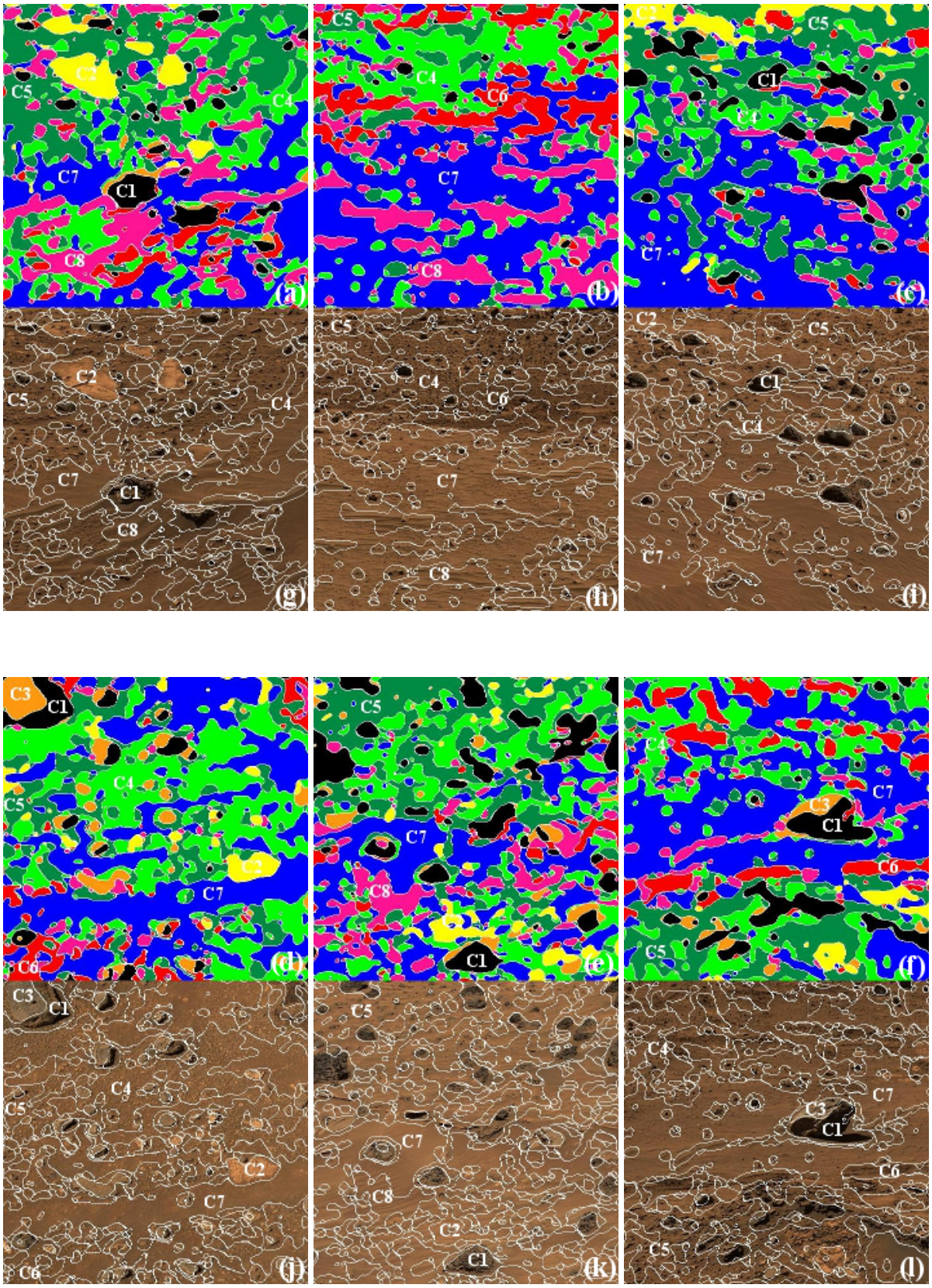

Fig. 5. Classified and segmented image.

retains the underlying semantics of the selected feature subset. This is very important to help ensure that the classification results are understandable by the user. Following this approach, the conventional SVM, KNN and DTREE, which are sensitive to the dimensionality of feature patterns, can be expected to become effective on classification of images whose pattern representation may otherwise involve a large number of features. Although the real-world images encountered are large-scale and complex, the resulting feature pattern dimensionality of selected features is manageable. In particular, SVM (and also KNN and DTREE) classifiers that are built using such selected features generally outperform those using more features or an equal number of features obtained by conventional dimensionality reduction techniques that are represented by PCA. This is confirmed by systematic experimental investigations. In short, this work helps to accomplish challenging image classification tasks effectively and efficiently. This is of particular significance for classification and analysis of real images on board or on ground in future Mars rover missions.

Acknowledgements

This work was partly supported by a Daphne Jackson Fellowship, funded by the Royal Academy of Engineering and partly funded by the European Community's 7th Framework Programme (FP7/2007-2013), Grant Agreement No. 218814 PRoVisG.

8. REFERENCES

[1] Castano R. *et al*. (2006). Current Results from a Rover Science Data Analysis system. *Proc. of IEEE Aerospace Conf*.

[2] http://marswatch.astro.cornell.edu/ pancam_instrument/mcmurdo_v2.html

[3] Woods M., Shaw A., Barnes D., Price D., Long D., Pullan D. (2009). Autonomous Science for an ExoMars Rover-like Mission. *Journal of Field Robotics*. 26(4), pp358-390.

[4] Kim W.S., Steele R.D., Ansar A.I., Al K., & Nesnas I. (2005). Rover-Based Visual Target Tracking Validation and Mission Infusion. *AIAA Space*. pp6717-6735

[5] Thompson D.R. & Castano R. (2007). Performance Comparison of Rock Detection Algorithms for Autonomous Planetary Geology. *Proc. of IEEE Aerospace Conf.*

[6] Shang C., Barnes D. & Shen Q. (2009). Taking Fuzzy-rough Application to Mars: Fuzzy-rough Feature Selection for Mars Terrain Image Classification. *Proc. of Int. Conf. on Rough Sets*. Springer. pp209-216.

[7] Huang K. & Aviyente S. (2008). Wavelet Feature Selection for Image Classification. *IEEE Trans. Image Proc*. 17, pp1709-1720.

[8] Kachanubal T. & Udomhunsakul S. (2008). Rock Textures Classification Based on Textural and Spectral Features. *Proc. of World Academy of Science, Eng. and Tech*. 29, pp110-116.

[9] Jensen, R & Shen Q. (2008). *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*. IEEE Press and Wiley.

[10] Puig D. & Garcia M.A. (2006). Automatic Texture Feature Selection for Image Pixel Classification. *Pattern Recognition*, 39, pp1996-2009.

[11] Jensen R. & Shen Q. (2009). New Approaches to Fuzzy-rough Feature Selection. *IEEE Trans. Fuzzy Syst*. 17(4), pp824-838.

[12] Vapnik, V. (1998). it Statistical Learning. New Yark, Wiley.

[13] Chapelle O., Haffner P. & Vapnik V.N. (1999). Support Vector Machines for Histogram-based Image Classification. *IEEE Tran. on Neural Networks*. 10, pp1055-1064.

[14] Mitchell T.M. (1997). *Machine Learning*. McGraw-Hill.

[15] Duda R.O., Hart P.E. & Stork D.G. (2001). *Pattern Classification*. (2nd edition). Wiley & Sons, New York.

[16] Martin D.R., Fowlkes C.C. & Malik J. (2004). Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. *IEEE Trans. Patt. Anal. and Mach. Inte*. 26, pp530-549.

[17] Gonzalez R.C. & Woods R. E. (2008) *Digital Image Processing and Scene Analysis* (3rd Edition), Prentice Hall.

[18] Cristianini N. & Shawe-Taylor J. (2000). An Introduction to Support Vector Machines and other Kernel-based Learning Methods. Cambridge University Press.

[19] Platt J. (1999). Macines Using Sequential Minimal Optimisation. in *Advances in Kernel Methods: Support Vector Learning*. MIT Press Cambridge, USA. pp185-208.